C00 Exercise 7: experiment

Ramana Kumar (rk436)

December 6, 2010

1 Introduction

One would expect T9 (predictive text) to be faster for text entry than multi-tap, since fewer keystrokes are required to input each word. However, the assumption of fewer T9 keystrokes can be violated by unpredicted words, which need to be entered in pieces. We test whether the (assumed) speed advantage also fails in this case.

Hypothesis For text with many unpredicted words, input by multi-tap is faster than by T9.

2 Method

2.1 Participants

We had $n = 10$ participants.

Demographics 5 were male, and 5 were female. Ages ranged from 22 to 26, with mean 24.4 and median 24.5.

Experience Participants ranged in experience with texting from none to ≥ 10 years, with the majority nominating 6–9 years. A majority, 7, indicated that they usually prefer T9, including 2 who indicated a strong preference.

2.2 Apparatus

We used the LG GS101 mobile phone, shown in Figure 1. The phone has 20 keys, whose functions during text entry are listed below.

Key(s)	Function
— (top left), green handset	not used
red handset	save to drafts and exit
up, down	cycle through T9 predictions
left, — (top right)	backspace
right	accept prediction/insert space
1	insert punctuation
2–9	insert letter
0	insert space
*	symbol input window, cycle through predictions
#	cycle through input methods

Figure 1: LG GS101 mobile phone

Each numeric key stands for a list of characters, for example 6 is [m, n, o, 6, ñ, ô, ö, œ] and 0 is [space, 0]. The initial part of the list is printed on the key; the rest is made up of related letters.

In T9 entry, after the user presses a sequence of numeric keys, the phone chooses a possible word with one character from each successive list. The possibilities and their priorities are taken from a dictionary. The user can cycle through alternative words, if necessary, to choose the right one. The dictionary may not contain every desired word.

In multi-tap entry, repeatedly pressing a numeric key cycles through the list of characters for that key. If 0 or right is pressed, or if no key is pressed after a short delay, the selection is inserted.

The # key cycles through input methods in the following order: [T9Abc, T9abc, T9ABC, 123, Abc, abc, ABC]. The first three are T9, and the last four are multi-tap. The suffixes determine the case of the next word: Abc indicates first-letter-only uppercase, abc indicates lowercase, and ABC indicates all uppercase. The method 123 is for entering numbers directly, and was not used.

For timing, we used an ordinary stopwatch.

2.3 Procedure

1. Ask participant to fill out consent form and questionnaire.
2. Show participant the text entry screen and cycle through the input methods once, stopping on T9Abc for T9-first participants, or Abc for multi-tap-first participants.
3. (For T9-first participants, otherwise after 6) Explain "If a word is not in the T9 dictionary, you may have to enter it in pieces."
4. Inform participant about task: "We will time how long it takes to enter the following text (shown on paper):

   'Left Stefan's beanie with MPhil stuff.'

   Punctuation and capitalisation are not important."
5. Time participant's text entry.
6. Say "Now we will time how long it takes to enter the same text with the other input method."
7. Clear the text window and cycle the input method to Abc for T9-first participants, or T9Abc otherwise.
8. Time participant's text entry.
9. Thank participant and ask for comments.

Consent form

This experiment concerns the speed of two common text-entry methods for mobile phones. You will be asked to enter the same text twice into a mobile phone, once for each input method, and you will be timed on a stopwatch. You will also be asked to fill in a short questionnaire about your prior texting experience, and for comments afterwards. Overall, the experiment should take 5–10 minutes.

No data collected during this experiment will be stored in a form that could be used to identify you. You may withdraw from the experiment at any time without penalty. In that case, all collected data will be destroyed. This experiment is without remuneration.

Questionnaire

1. What is your age? _____
2. What is your sex?
   Male, Female
3. How much experience do you have using mobile phones for texting?
   None, A few months, 1–5 years, 6–9 years, $\geq$ 10 years
4. How often do you use T9 (predictive text) for text entry?
   Never, Sometimes, Most of the time, Always
5. How often do you use multi-tap for text entry?
   Never, Sometimes, Most of the time, Always
6. Which do you prefer out of T9 and multi-tap?
   Strongly prefer multi-tap, Prefer multi-tap, Don't know, Prefer T9, Strongly prefer T9

2.4 Design

We tested two factors: T9 input and multi-tap input. We used one task: entering a sentence. Thus there were two conditions.

We used a repeated measures design: each participant was exposed to both conditions. For counterbalancing, half the participants did T9 entry first, and the other half did multi-tap entry first. Thus the participants are divided into two groups: T9-first and multi-tap-first.

The test sentence contains 6 words, of which 3–"Stefan's", "beanie", and "MPhil"–are not in the T9 dictionary. It was chosen to have the following attributes:

- reasonably plausible
- at least half the words unpredicted
- makes use of all the numeric keys
- can take less than a minute to enter

3 Results

Data and R script file for the results are available at `http://www.cl.cam.ac.uk/~rk436`.

Entering the test sentence "Left Stefan's beanie in MPhil lab.", which contains 3 words that are unpredicted by T9, took longer using T9 than it took using multi-tap. The mean time for T9 entry was 121.4 seconds, and for multi-tap entry was 56.2 seconds. A box plot summarising the times is given in Figure 2.

An independent samples t test showed a significant effect of the multi-tap method against the T9 method on the time for text entry ($t(9.7) = 2.56, p < 0.05$).

The time for T9 entry can be explained by T9 usage, with significance at the 95% level ($p < 0.05$). The relationship between T9 times and T9 usage is shown in Figure 3. By contrast, there was no significant relationship between multi-tap times and multi-tap usage (Figure 4).

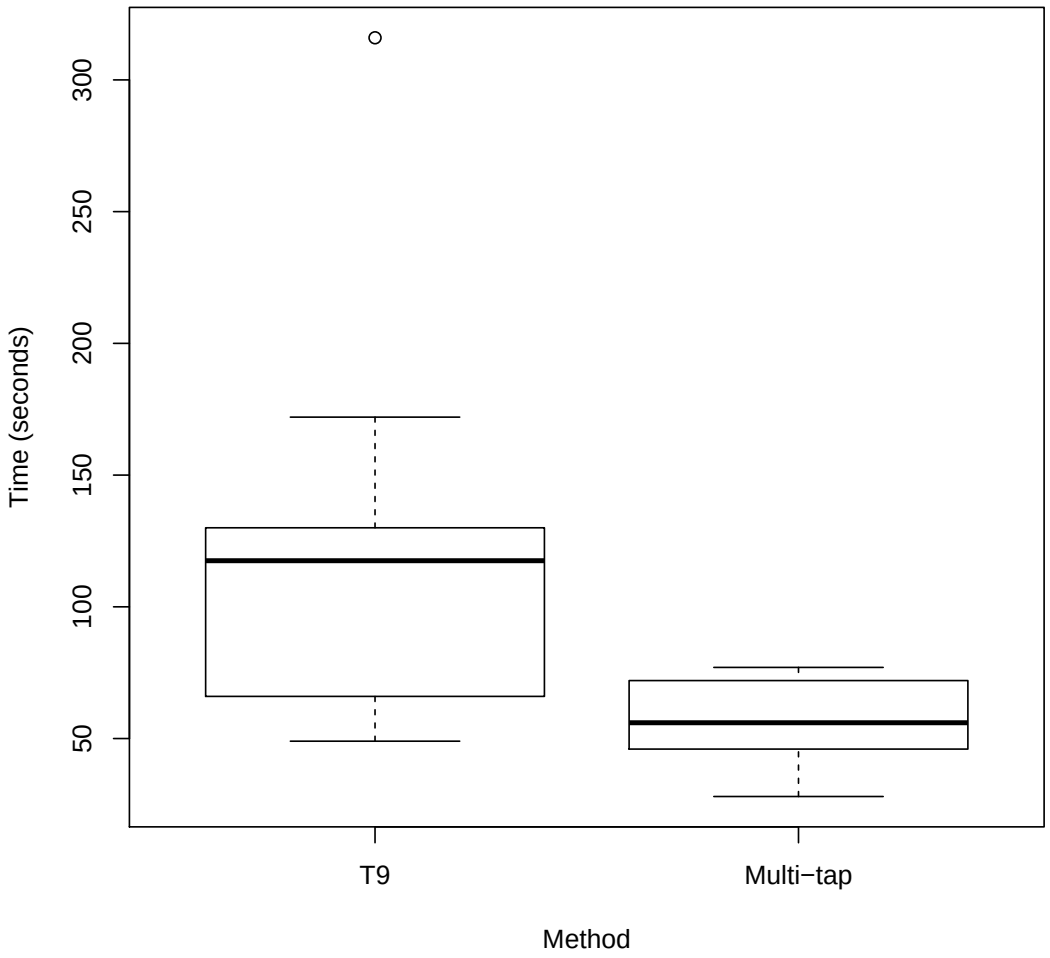

Figure 2: Time taken to enter 6 words, where 3 words are unpredicted by the T9 dictionary. The 10 participants were each timed once using T9 and once using multi-tap (half did T9 first). The mean time for T9 was 121.4 s, and for multi-tap was 56.2 s. The standard deviation for T9 was 79.0 s, and for multi-tap was 15.1 s. The outlying point is due to a participant with no prior texting experience; the same participant had less trouble with multi-tap entry.

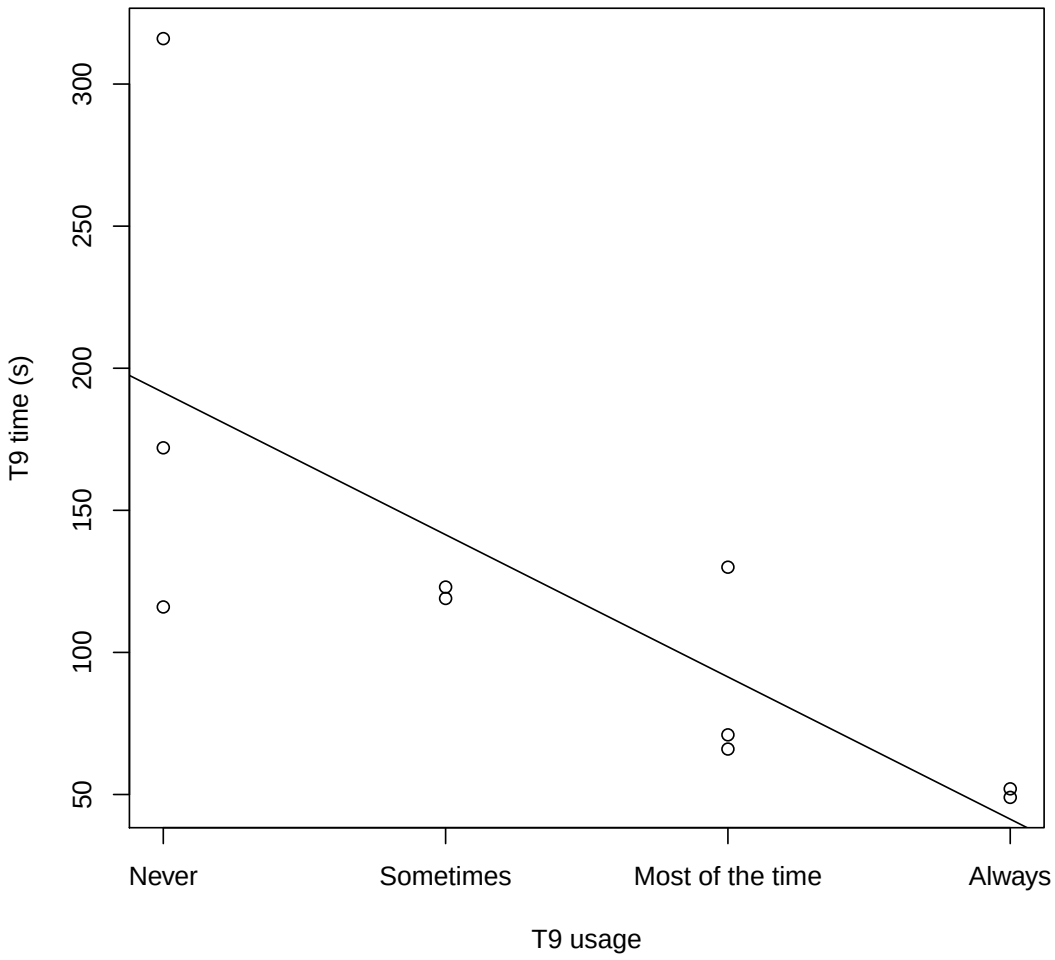

Figure 3: The time taken to enter the test sentence using T9 depends on the participant's familiarity with T9. The linear model shown has a p-value of 0.01374.

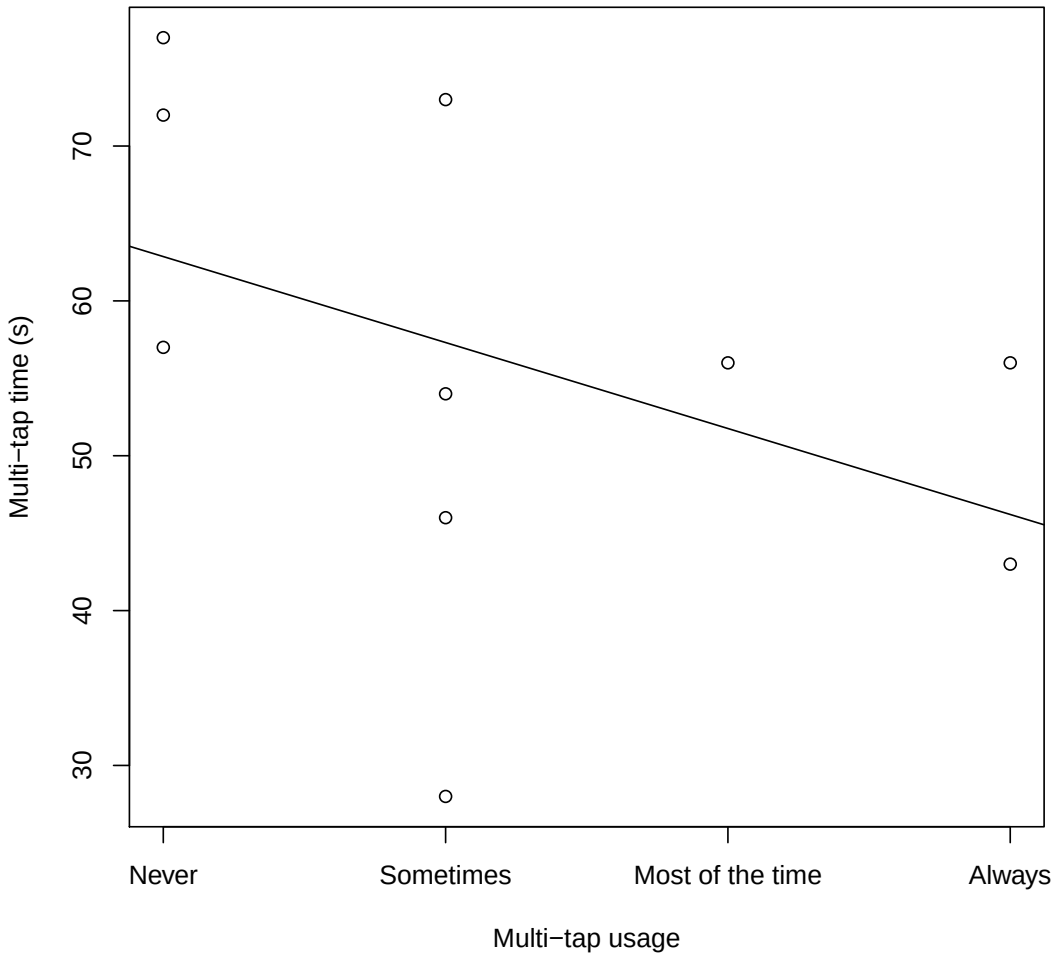

Figure 4: There is no significant relationship between the time taken to enter the test sentence using multi-tap and the participant's familiarity with multi-tap. The fitted linear model has a p-value of 0.229.

There was no significant learning effect during the experiment, although participants in the multi-tap-first group were on average faster. The mean time for T9 entry was 157.2 s in the T9-first group, and 85.6 s in the multi-tap-first group (Figure 5). An independent samples t test showed no significant difference between groups on T9 time ($t(4.82) = 1.54, p > 0.05$). The mean time for multi-tap entry was 58.0 s in the T9-first group, and 54.4 s in the multi-tap-first group (Figure 6). An independent samples t test showed no significant difference between groups on multi-tap time ($t(6.52) = 0.359, p > 0.1$).

4 Discussion

Our results support the hypothesis presented in the introduction, that T9 is slower than multi-tap for text with many unpredicted words. If we assume T9 is faster for predicted words, we are led to the conclusion that effective text entry involves choosing the right input method for the task at hand. Indeed, more than one participant commented that the two methods are good for different purposes, for example multi-tap is better for names and other unusual words. Some participants said they would usually switch between methods as necessary.

An hypothesis worth further investigation is whether T9 is more difficult to learn or more unintuitive than multi-tap, although it may be more effective for experienced users. Our results show that T9 time depends significantly on prior experience with the method, whereas multi-tap time does not depend so much on experience. We also see a greater variation in T9 times. Informally, we observed that some participants, who were less familiar with T9, treated the method like multi-tap by entering and correcting one letter at a time rather than syllable-length chunks at a time. Perhaps users under time pressure easily lose faith in the predictive ability of T9.

In this experiment we explicitly ignored punctuation and capitalisation in the test sentence. This may be more in keeping with common texting behaviour. Indeed, no participants entered the test sentence with perfect punctuation and case. Errors included using all capitals, multiple spaces (sometimes within words), and missing punctuation. It would be interesting to see whether enforcing corrections of these errors increases or decreases the gap between multi-tap and T9.

Some participants commented that the T9 method on their own phones was different than on the test phone. For example, the test phone inserts a space automatically after you accept a word using the right arrow key. Familiarity with T9, then, can be specific to a particular model, and may be a disadvantage for other models. There is probably less variation in multi-tap.

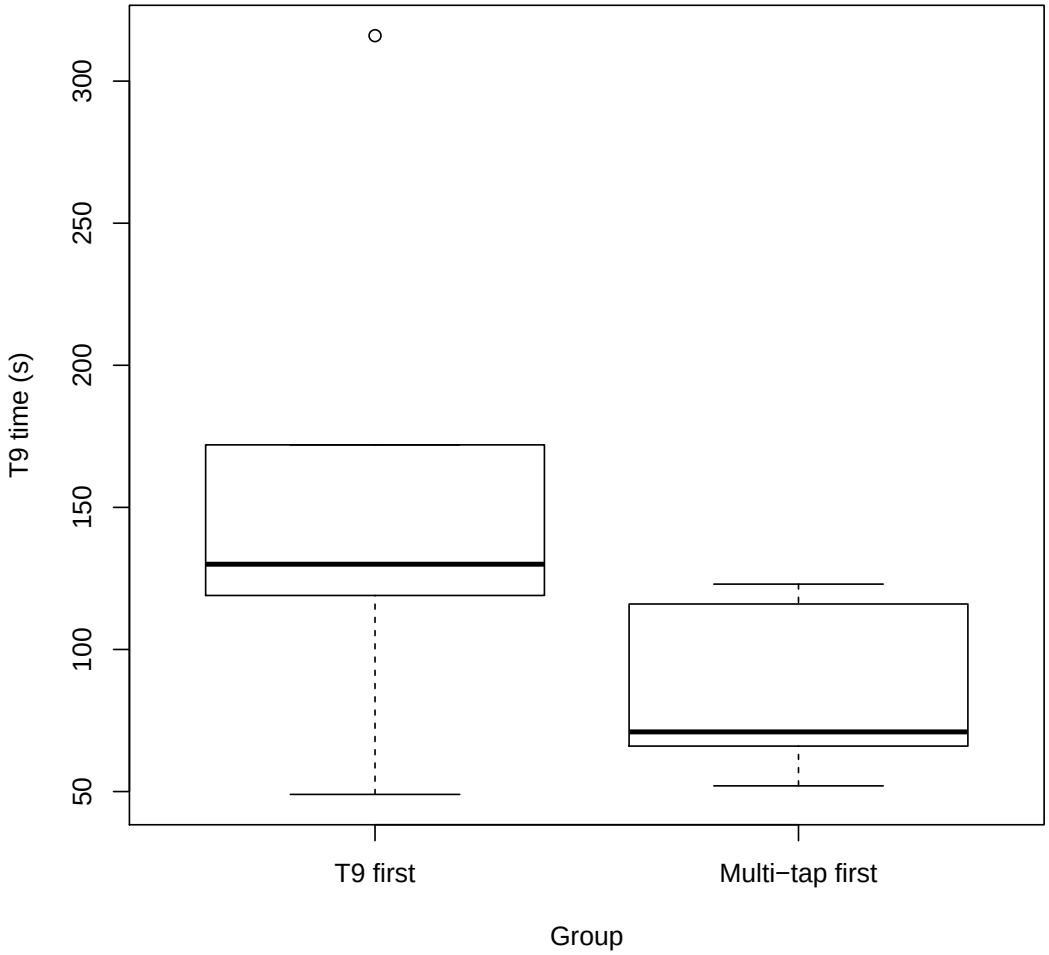

Figure 5: For the time taken to enter the test sentence using T9, there was no significant difference between the 5 participants who did T9 first and the 5 who did multi-tap first. The p-value here for an independent samples t test is 0.0935.

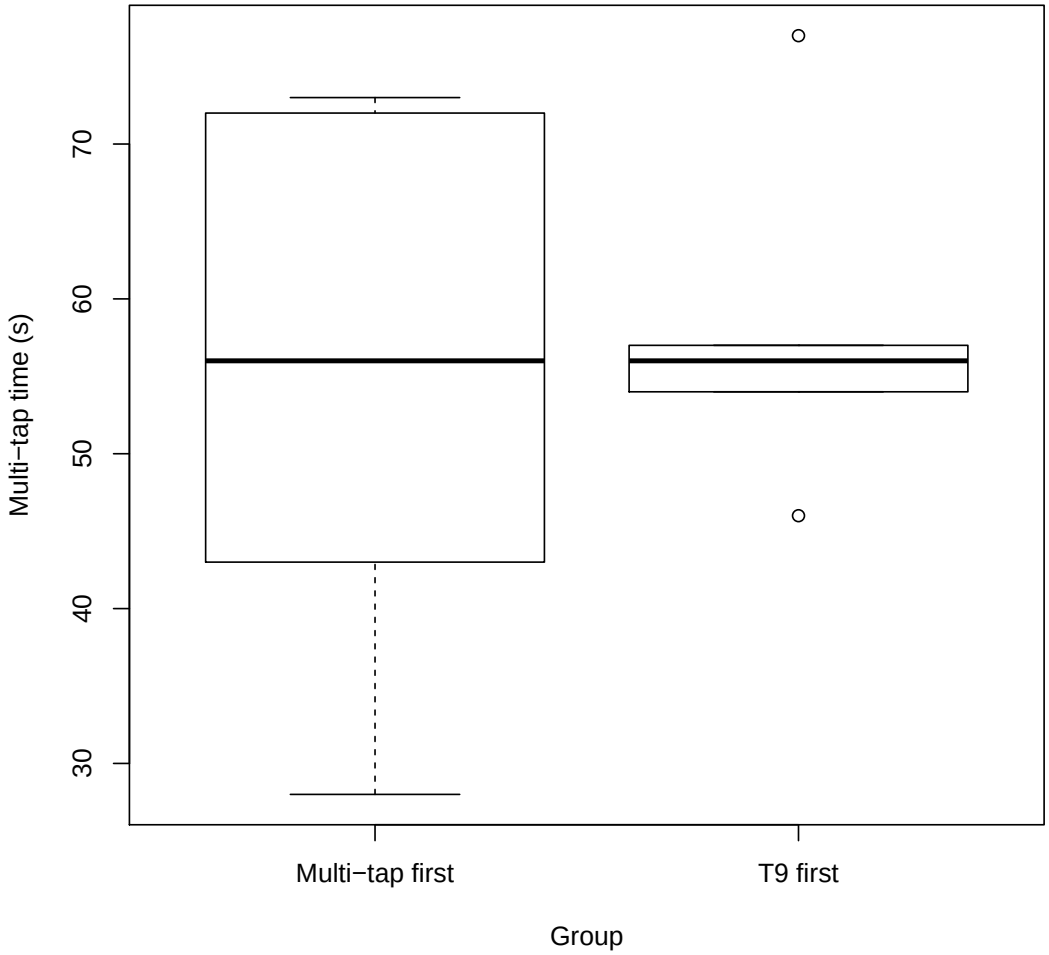

Figure 6: For the time taken to enter the test sentence using multi-tap, there was no significant difference between the 5 participants who did multi-tap first and the 5 who did T9 first. The p-value here for an independent samples t test is 0.366.